

## ANALISIS CURAH HUJAN DAN TEMPERATUR RATA-RATA PADA DATA METEOROLOGI BMKG KOTA JAMBI MENGUNAKAN METODE CLUSTERING K-MEANS

**Muhammad Yana Haris Farsya**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara  
Jl. Letjen S. Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410  
Email: muhammad.535190036@stu.untar.ac.id

### ABSTRAK

Suhu udara adalah kondisi udara pada waktu dan tempat tertentu. Jika diperhatikan dengan seksama, bahwa suhu udara di berbagai tempat berbeda. Ada yang suhunya terlalu dingin, terlalu panas, dan ada pula yang seimbangan. Dengan mengetahui suhu rata-rata, maka curah dapat di perkirakan. Curah hujan adalah suatu kejadian alam yang sangat berpengaruh dari segi produktivitas dan juga aktivitas lainnya. Hal negatif dari hujan adalah jika seandainya seseorang berangkat ke kantor lalu terjadi hujan besar maka perjalanan akan terlambat, atau jika sedang menyuci baju maka baju akan basah dikarenakan hujan. Sedangkan untuk hal positifnya dengan adanya hujan maka daerah yang kering bisa terselamatkan atau dapat juga membantu para petani agar ladang tidak kekeringan. Rancangan ini ditujukan agar dapat melihat statistik clusternya apakah di daerah Jambi terdapat curah hujan serta temperatur rata-rata yang cukup atau tidak. Rancangan dibuat menggunakan metode clustering K-means dengan variabel yang digunakan adalah RR dan juga Tavg.

**Kata kunci:** Curah Hujan, Temperatur, Klasterisasi, *K-Means*, Meteorologi

### ABSTRACT

*Air temperature is the condition of the air at a specific time and place. If you pay close attention, you will notice that the air temperature varies from place to place. Some places are too cold, some are too hot, and some have a balanced temperature. By knowing the average temperature, rainfall can be estimated. Rainfall is a natural phenomenon that greatly affects productivity and other activities. The negative aspect of rain is that if someone leaves for work and it rains heavily, they will be late, or if they are washing clothes, the clothes will get wet because of the rain. On the positive side, rain can save dry areas or help farmers prevent their fields from drying up. This design aims to see whether the cluster statistics show that the Jambi area has sufficient rainfall and average temperature. The design was created using the K-means clustering method with the variables RR and Tavg.*

**Keywords:** Rainfall, Temperature, Clustering, *K-Means*, Meteorology

## 1. PENDAHULUAN

Cuaca merupakan hal yang paling penting pada kehidupan manusia. Cuaca adalah keadaan udara pada saat tertentu dan di wilayah tertentu yang relatif sempit dan pada jangka waktu yang singkat.[1] Perubahan cuaca akan menyebabkan terganggunya kegiatan produksi maupun distribusi dalam melakukan sesuatu sehingga dibutuhkan analisis yang tepat dalam menanggapi cuaca. Pada rancangan ini akan dilakukan *clustering* pada temperature minimum dan curah hujan berdasarkan data yang ada. Proses terjadinya hujan berhubungan dengan unsur-unsur cuaca lainnya seperti temperatur udara, kelembaban udara, kecepatan angin, dan arah angin.[2] Maka dari itu rancangan akan mengambil data yang, yaitu curah hujan dan temperatur rata-rata. Rancangan akan menggunakan metode Clustering K-Means pada data BMKG kota Jambi.

Tujuan dari menganalisis data BMKG kota Jambi adalah untuk menentukan apakah curah hujan dan temperatur rata-ratanya cukup atau tidak. Dengan hasil yang akan diambil, jika seandainya curah hujan dan temperature berada di tingkat rendah maka akan diusahakan agar wilayah tersebut

tidak kekeringan dengan hujan buatan atau hal lainnya. Dengan *clustering* ini, diharapkan agar dapat membantu banyak orang dalam melakukan berbagai kegiatan tanpa terganggu oleh kendala yang akan datang dan juga bisa mempersiapkan diri untuk tetap melakukan kegiatannya.

## 2. METODE

Rancangan akan dibuat menggunakan metode Clustering K-Means. Algoritma K-Means Clustering sebagai salah satu metode yang mempartisi data ke dalam bentuk satu atau lebih cluster atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Kelompok atau cluster yang didapat merupakan pengetahuan/informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan.[3]

Teknik evaluasi yang digunakan adalah Silhouette Coefficient. Metode Silhouette Coefficient merupakan gabungan dari metode *cohesion* dan *separation*. Metode ini digunakan untuk melihat kualitas dan kekuatan cluster yaitu seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Selain itu dapat juga digunakan untuk mengukur seberapa dekat relasi antar objek dalam sebuah *cluster*. Metode *separation* yang berfungsi untuk mengukur seberapa jauh sebuah *cluster* terpisah dengan *cluster* lain.

Tahapan perhitungan Silhouette Coefficient adalah sebagai berikut[5]:

A. Hitung rata-rata jarak dari suatu objek dengan persamaan (1).

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (1)$$

B. Hitung rata-rata jarak dari objek I dengan semua objek dari cluster lain seperti dengan persamaan (2).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

C. Nilai Silhouette Coefficient ditentukan menggunakan persamaan (3).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data

Dataset bersumber dari BMKG yang diperoleh dari situs <https://dataonline.bmkg.go.id/> yang diambil mulai tahun Januari 2010 sampai Oktober 2022 yang memiliki 10 variabel sebagai berikut:

1. Tn = Temperatur minimum
2. Tx = Temperatur maximum
3. Tavg = Temperatur rata-rata
4. RH\_avg = Kelembapan rata-rata
5. RR = Curah hujan
6. ss = Lamanya penyinaran matahari
7. ff\_x = Kecepatan angin maksimum
8. ddd\_x = Arah angin saat kecepatan maksimum
9. ff\_avg = Kecepatan angin rata-rata
10. ddd\_car = Arah angin terbanyak

### 3.2 Pra-Pemrosesan Data

Pada tahap Pra-pemrosesan data, terdapat beberapa *missing value* di beberapa variabel, *missing value* akan diisi dengan berdasarkan persentase *missing value* tiap bulannya. Jika dalam 1 bulannya terdapat *missing value* kurang dari 20% maka data yang bernilai numerik diisi dengan nilai rata-rata dari 7 hari sebelumnya, untuk yang *non* numerik, maka data akan diisi dengan nilai modulus dari 7 hari terakhir. Sedangkan jika *missing value* pada 1 bulan lebih dari 20% maka untuk nilai numerik akan diisi dengan nilai rata-rata dari data di bulan yang sama dari tahun-tahun sebelumnya, sedangkan untuk nilai yang *non* numerik akan diisi dengan nilai modulus dari data di bulan yang sama dari tahun sebelumnya. Berikut adalah jumlah persentase *missing value* yang ada pada tiap variabel:

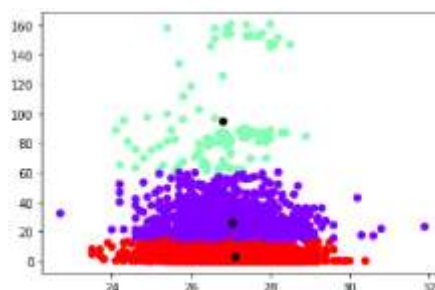
1. Tn (Temperatur minimum)	= 1,051965657%
2. Tx (Temperatur maximum)	= 1,032830524%
3. Tavg (Temperatur rata-rata)	= 1,00605013%
4. RH_avg (Kelembapan rata-rata)	= 1,00626756%
5. RR (Curah hujan)	= 1,166917293%
6. ss (Lamanya penyinaran matahari)	= 1,058181818%
7. ff_x (Kecepatan angin maksimum)	= 1,004314064%
8. ddd_x (Arah angin saat kecepatan maksimum)	= 1,004314064%
9. ff_avg (Kecepatan angin rata-rata)	= 1,004530744%
10. ddd_car (Arah angin terbanyak)	= 1,004314064%
Keseluruhan <i>missing value</i>	= 10,33968592%

Berdasarkan data di atas, tiap variabel terdapat *missing value* 1% dengan ke seluruh *missing value* adalah 10%. Hasil akhir dari pra-processing adalah mengisi data-data yang kosong berdasarkan ketentuan yang sudah dibuat.

### 3.3 Eksperimen

Pada dataset BMKG yang terdapat 10 variabel, rancangan akan melakukan *clustering* pada variabel RR dan juga Tavg. Curah hujan adalah jumlah curah hujan yang dinyatakan dalam tinggi hujan atau volume hujan tiap satuan waktu, yang terjadi pada satu kurun waktu air hujan terkonsentrasi. Besarnya intensitas curah hujan berbeda-beda tergantung dari lamanya curah hujan.[5]

Pada Rancangan akan melakukan *clustering* pada temperatur rata-rata dan juga curah hujan pada tahun 2010 Januari sampai dengan 2022 Desember dengan jumlah *cluster* yang akan dibagi menjadi 3. Berikut merupakan visualisasi dari hasil *clustering* 3 bagian pada data temperatur rata-rata dan juga curah hujan.



**Gambar 1.** Visualisasi clustering berdasarkan data 2010 sampai dengan 2022.

### 3.4 Evaluasi

Evaluasi dihasilkan menggunakan Silhouette Coefficient yang merupakan pendekatan dari 1 *cluster* ke *cluster* lain serta tiap objek di dalam 1 *cluster*. Hasil dari rata-rata nilai silhouette adalah seperti pada gambar 2.

Rata - rata nilai Silhouette = 0.7033642922932506

**Gambar 2.** Hasil nilai silhouette

### 3.5 Analisis

Berdasarkan hasil clustering yang ada pada gambar 1, dapat dinyatakan bahwa pada cluster bewarna hijau terang memiliki curah hujan yang tinggi dengan temperatur yang juga tinggi, sedangkan pada cluster ungu berada di tingkat lumayan sedangkan pada cluster merah berada di tingkat rendah. Dengan hasil silhouette rata-rata dari *cluster* tersebut adalah 0.703 dimana nilainya mendekati 1 yang hasilnya berarti *clustering* kurang bagus.

## 4. KESIMPULAN

Kesimpulan yang didapat pada hasil analisi data BMKG kota Jambi adalah nilai silhouette pada *cluster* cukup buruk sehingga hasil *cluster* yang didapatkan kurang akurat. Saran yang dapat diberikan adalah untuk penelitian kedepannya dapat menggunakan metode lain sebagai acuan apakah hasil dari Clustering K-Means bagus atau tidak.

## DAFTAR PUSTAKA

- [1] Badan Meteorologi dan Geofisika, "Cuaca dan Iklim Maritim," vol. 02, p. xi+85, 2018, [Online]. Available: <https://maritim.bmkg.go.id/bulletins/2018/pdf/12-Desember.pdf>
- [2] S. F. Rohmana, A. Rusgiyono, and S. Sugito, "Penentuan Faktor-Faktor Yang Mempengaruhi Intensitas Curah Hujan Dengan Analisis Diskriminan Ganda Dan Regresi Logistik Multinomial (Studi Kasus: Data Curah Hujan Kota Semarang dari Stasiun Meteorologi Maritim Tanjung Emas Periode Oktober 2018 – Maret 2019)," *J. Gaussian*, vol. 8, no. 3, pp. 398–406, 2019, doi: 10.14710/j.gauss.v8i3.26684.
- [3] A. Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>
- [4] Juleha, Rismalinda, and A. Rahmi, "Analisa Metode Intensitas Hujan Pada Stasiun Hujan Rokan Iv Koto, Ujung Batu, Dan Tandun Mewakili Ketersediaan Air Di Sungai Rokan," *J. Mhs. Tek. UPP*, vol. 1, no. 1, pp. 1–8, 2016.